Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 10 (2017), pp. 7669–7688

http://www.ripublication.com/gjpam.htm

A comparison of three different enhancements of the generalized estimating equations method in handling incomplete longitudinal binary outcome

Aluko Omololu S. and Mwambi H.

*School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Private Bag X01 Scottsville 3209,
Pietermaritzburg, South Africa.*

Abstract

This paper compares the performance of three techniques of analyzing incomplete longitudinal binary outcome data when the missingness is due to dropout. It is assumed the response data are missing at random. We consider three modifications of the generalized estimating equations (GEE) based on inverse probability weighting (IPW) and multiple imputation (MI). In the weighted GEE (WGEE), we weight observations by the inverse of the probability of being observed. The multiple imputation (MI) combined with GEE analysis is commonly known as MI-GEE. In this approach, the missing observations are filled multiple times with the predicted values from the imputation model followed by a GEE analysis. The so-called doubly-robust (DR) technique combines the multiply imputed binary responses with IPW and then applying GEE to the completed data sets. A simulation study is first used to compare the performance of the methods followed by an application to a clinical trial data on Amenorrhea. The simulation and empirical example results revealed better performance for DR-GEE compared to WGEE and MI-GEE, but MI-GEE was evidently superior than WGEE and quite close to DR-GEE.

AMS subject classification:
Keywords: Missing at Random, Multiple Imputation, Inverse probability weighting, Doubly Robust.

1. Introduction

In a longitudinal study, each individual or unit is measured at several time points which provides the opportunity to study changes over time for the variable(s) and effects of interest. In several areas of biomedical applications the response variable is binary or in general non-Gaussian, for instance, the presence or absence of an ailment in an individual included in a clinical trial to compare two or more treatments or interventions. The most widely used approach for handling binary longitudinal responses is the generalized linear mixed model (GLMM) [7, 17]. In the presence of missing data this model imply complex and hard to manipulate likelihoods for moderate and large sequences of repeated measurements. Generalized estimating equations (GEE) are on alternative modeling technique [14], but needs some enhancements to deal with incomplete longitudinal data where missing completely at random (MCAR) is ruled out. The essence of this technique is that, it allows confining attention to the mean structure given that there is a willingness to adopt a working assumption about the association structure for the repeated measurements.

Thus, the main difference is that the GLMM is a conditional or random effects likelihood based model while the former namely the GEE is a marginal model. The strength of the GLMM is that it is likelihood based and hence in the presence of missing data the obvious missing at random (MAR) assumption can be naturally accommodated.

In the presence of incomplete data, GEE suffers from its frequentist nature where its validity is restricted to the MCAR assumption, meaning the missingness is independent of both unobserved and observed responses [19]. To overcome this defficiency, another member of the GEE technique was introduced by [22]. This is the WGEE because it allows for the weaker MAR assumption, where the missingness is independent of the unobserved data given the observed data [1, 19]. Under WGEE, the technique uses the inverse of the individual's probability of being observed as a weight to the estimating equations in order to reduce bias in the regression parameter estimates.

The GEE method is one of the most common techniques for the analysis of non-Gaussian correlated data. It is advantageous when one specifies the mean structure correctly for the parameter estimates to be consistent and asymptotically normal. In deriving the method, the association parameter(s) among the repeated measures are taken as nuisance parameters. The GEE technique is attractive because it helps avoid dealing with complex and sometimes intractable likelihoods and naturally the interpretation of the parameters of interest are population-averaged. When longitudinal data is incomplete, the non-response can occur at any time from the beginning of the study. There are two patterns of missing data that can be observed for the response: first, dropout is when an individual leaves the study prematurely for reasons known or not known to the investigator and does not return. This generally falls under the monotone pattern of non-response. Second, intermittent non-response occurs when an individual leaves and returns to the study after some period of non-response, and possibly a repeat of the same once, twice or more times. Missing data is also possible in the covariates, but our focus in this paper is on missing data in the outcome variable. In the presence of incomplete data, three challenges are of concern. First, between the observed and missing data there can be potential serious bias due to systematic differences. Second, handling the data and statistical inferences are complex and lastly, the loss of efficiency can be substantial. Furthermore, in order to have inferences that are valid; it is pertinent to have a good knowledge of the missing data mechanisms that could have generated the non-response and how to handle it.

Doubly robust estimators (DR-GEE) are seen as an appealing modification, or extension of the ordinary GEE to handle data that are subject to MAR mechanism. The doubly robust (DR) estimating equations method has been developed as an extension of the WGEE method, where the idea is to integrate the weights with the use of a predictive imputation model for the missing data given the observed data. In effect, the DR estimation method produces parameter estimates that are consistently correct given correct specification of either the weights or the predictive imputation model, but not necessarily both.

DR techniques have widely received attention in the literature in the last decade (see [1, 3, 12, 13, 28]). More importantly, the inclusion of the inverse of the propensity score into the imputation model gives an increasing robustness to the imputations against misspecification of the imputation model. The uniqueness of this technique is that it gives analysts two routes to validate inferences, instead of only one. However, the method lacks generalization to intermittent missing observations, where the individuals return to the study after skipping one or more visits. Multiple imputation (MI) is one of the alternative approaches [15] which relies on the MAR assumption. Under this approach missing values are imputed several times, and the resulting completed datasets are analyzed using a standard technique like GEE. Several authors like [2] have worked on the combination of MI and GEE, such that missing data are multiply imputed, and then inferences are obtained based on GEE. These inferences are combined into a single summary using Rubin's pooling rules and hence the method has become commonly known as MI-GEE. Moreover, the important requirements in this method are just like any other technique for imputation, namely the imputation model needs to be specified correctly. That the model should include all important covariates; including auxiliary ones to make it rich in informing the missing values predictive process. By its very nature MI-GEE does not suffer from the intermittent missing data problem.

A study conducted by [24], compared the two types of GEE (WGEE and MI-GEE) to the likelihood-based GLMM for analyzing longitudinal binary outcomes with dropout. The use of extended or enhanced GEEs to other categorical outcomes has and is also gaining popularity. For example, authors in [26] used a WGEE method to accommodate arbitrary patterns of a missing responses and missingness in key covariates. A recent paper from [5], compared through a simulation study two multiple imputation methods (multivariate normal imputation and ordinal imputation regression) for longitudinal ordinal data subject to dropout. In another paper, the same authors compared joint modeling and fully conditional specification approaches for non-monotone [6]. Single robust versions of GEE are used in the two papers mentioned above and they treated only a MAR baseline covariates or MAR response. In a recent paper, [13] proposed a doubly robust approach for the analysis of longitudinal ordinal data with intermittently missing response and covariates under MAR.

In this paper, we re-visit the incomplete binary data problem. Our interest is thus on the combination of the DR and GEE for incomplete longitudinal binary data when the missing data pattern is monotone. This method involves multiply imputing binary responses using the DR approach and then applying GEE to the complete data sets. However, we also introduce a novel idea to find out whether using different working correlation structures, namely; compound symmetry (CS), first order autoregressive AR(1) and TOEP for estimation would affect the parameter estimates and standard errors under the specified methods. This paper is organized as follows. Section 2 defines the GEE notations and an overview of the GEE method. Section 3, outlines the WGEE, MI-GEE and DR-GEE approaches. A simulation study is presented in section 4. Data arising from Amenorrhea study reported in [7] is analyzed in section 5. The paper ends with a discussion and conclusion in section 6.

2. The generalized estimating equation (GEE)

In the situation where population-averaged effects are of interest, the most widely used model to analyzing discrete longitudinal data are the GEE which falls among the popular marginal models. The GEE technique was proposed by [14], as an extension of the generalized linear model (GLM) to the case of correlated data in the context of longitudinal studies.

Suppose y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, N$, represents the jth response at say time t_{ij} on the ith individual with a vector of covariates x_{ij}. Thus n_i are the measurements on individual i, and let n be the maximum number of measurements per individual, i.e. $n = max_i\{n_i\}$ if all the planned repeated measurements were obtained. This is the most general setting but if $t_{ij} = t_j$ and $n_i = n$ for all i, then we have the most balanced design in the presence of no missing values.

We assume that the responses on the ith individual are held in the vector $Y_i = [y_{i1}, \ldots, y_{in_i}]'$ and the corresponding vector of means is $\mu_i = [\mu_{i1}, \ldots, \mu_{in_i}]'$. Under the generalized linear model formulation, the marginal mean μ_{ij} of the response y_{ij} is related to a linear predictor through a link function $g(\mu_{ij}) = x_{ij}^{'}\beta$ or as others may prefer $\mu_{ij} = h(x_{ij}^{'}\beta)$ where $h = g^{-1}$, and the variance of y_{ij} depends on the mean through a variance function $v(\mu_{ij})$, since $\text{var}(y_{ij}) = a(\phi)v(\mu_{ij})$ given $a(\phi)$ is the additional overdispersion parameter in the exponential family formulation. However, since we have repeated measurements the GLM has to be modified to account for the correlation of observations within an individual. This leads to a modified estimating equation for the model parameters. The generalized estimating equation, used to estimate the parameters of interest in the vector β in the marginal model is given by

$$S(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i'}{\partial \beta} \mathbf{V}_i^{-1} (Y_i - \mu_i(\beta)) = 0 \qquad (1)$$

where $\mathbf{V}_i$ is the covariance matrix of Y_i which is specified through the working correlation matrix $R_i(\alpha)$ as

$$\mathbf{V}_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2} \tag{2}$$

Here, thA_i is an $n_i \times n_i$ diagonal matrix whose jth diagonal element is $v(\mu_{ij})$, the variance function at μ_{ij}, from the assumed linear exponential family distribution. If $R_i(\alpha)$ is the true correlation matrix of Y_i, then V_i will be the true covariance matrix of Y_i and in this case the resulting standard errors of parameters are referred to as model based standard errors. Otherwise, it is suffices to use the empirical robust standard errors because in reality it is hard to discern the true covariance structure.

In the GEE estimation technique, the only requirements are the specification of the mean model and correlation structure of the vector Y_i; such that the specification of the full joint distribution of the correlated responses is not needed. The joint marginal distribution is complex to specify because often for non-Gaussian longitudinal responses, the joint distribution involves high-order associations and integration. However, the regression parameter estimates from the GEE are consistent even when the working covariance specification through $R_i(\alpha)$ is incorrect. When the marginal effects are of interest and the responses are not continuous, the GEE is a very common choice. Nevertheless, the GEE approach can lead to biased estimates when the underlying missingness mechanism is not MCAR. One of the methods that can produce unbiased estimates is the WGEE and is briefly described in the following section.

3. Methods for handling incomplete data

The weighted generalized estimating equations and multiple imputation are the two commonly used methods for missing data under the MAR mechanism. The missingness mechanism can be described via a statistical model for the probability of observing a missing value. A reasonable assumption about the mechanism is important for methods that are used to handle missing data. In general, missingness mechanisms are classified into three types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [19]. The three mechanisms are briefly discussed below in the contest of longitudinal data. We confine attention to dropout as the missing data pattern. First for each potential out Y_{ij} define a binary indicator variable R_{ij} which takes the value $r_{ij} = 0$ if Y_{ij} is missing and $r_{ij} = 1$ if Y_{ij} is observed.

- A missingness mechanism is said to be MCAR if the probability of a missing response is independent of its past, current and future responses conditional on the covariates. That is $P(r_{ij} = 0|Y_i, X_i) = P(r_{ij} = 0|X_i)$.
- A missing mechanism is said to be MAR if the probability of a missing response is independent of its current and future responses conditional on the observed past responses and the covariates. That is
  $P(r_{ij} = 0|r_{ij-1} = 1, X_i, Y_i) = P(r_{ij} = 0|r_{ij-1} = 1, X_i, y_{i1}, \dots, y_{ij-1})$
  MAR is a weaker assumption than MCAR. In fact, MCAR is a special case of MAR, thus an analyst is better of working with the superior MAR than the MCAR assumption.

- A missing mechanism is said to be MNAR if the probability of a missing response depends on the unobserved responses. MNAR is the most general and the most complex missing data mechanism to deal with. Thus, there is no further reduction to $P(r_{ij} = 0|r_{ij-1} = 1, X_i, Y_i)$.

Sections 3.1 to 3.3 present a brief discussion of the missing data methods considered in the current paper.

3.1. The weighted generalized estimating equation (WGEE)

Currently, there are two weighting methods that can be used to construct the WGEE for estimating the regression parameter β, when dropout is the missing data pattern. These are observation-specific weights and subject-specific weights versions as outlined in [10]. The weighting methods produce parameter estimates that are consistent provided the data are MAR.

3.1.1 WGEE based on observation-specific weights

The weight ω_{ij} for y_{ij} is defined as the inverse probability of observing y_{ij}. In other words, $\omega_{ij} = P(r_{ij} = 1|X_i, h_{ij})^{-1}$ where $h_{ij} = (y_{i1}, \ldots, y_{ij-1})$ denotes the observed response history. Let W_i be a $n_i \times n_i$ diagonal matrix whose jth diagonal is $r_{ij}\omega_{ij}$. Then the weighted generalized estimating equation [18, 23] is given by

$$S_{ow}(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} W_i (Y_i - \mu_i(\beta)) = 0 \tag{3}$$

Unlike the standard GEE, the weighted estimating equation is unbiased when observations are appropriately weighted, leading to consistent parameter estimates of β. Practically, the weights ω_{ij} are unknown and they have to be estimated using an appropriate model for r_{ij}, such as the logistic regression model under the MAR assumption. Specifically, suppose $\lambda_{ij} = P(r_{ij} = 1|r_{ij-1}, X_i, h_{ij})$ denote the probability of observing the response y_{ij} given previous observed responses, under the MAR assumption. Using the observed data, λ_{ij} can be predicted from the logistic regression model, logit $(\lambda_{ij}) = z_{ij}'\alpha$, where z_{ij} are predictors that usually include the covariates x_{ij}, the past responses and indicators for visit times and α is a vector of parameters. The dropout process implies that the estimated probability of observing y_{ij} can be expressed as a cumulative product of conditional probabilities given by

$\hat{P}(r_{ij} = 1|X_i, h_{ij}) = \lambda_{i1}(\hat{\alpha}) \times \cdots \times \lambda_{ij}(\hat{\alpha})$

With the estimated weights $\hat{\omega}_{ij} = \hat{P}(r_{ij} = 1|X_i, h_{ij})^{-1}$, we solve the estimating equation $S_{ow}(\beta)$, from which the regression parameter β is estimated. There is a similarity in the algorithm for solving the WGEE and standard GEE.

When the dropout process is MAR; the following algorithm fits marginal models by using the observation-specific WGEE method. The steps are:

S1. Fit a logistic regression with data (r_{ij}, z_{ij}) to obtain an estimate of α and estimate the weights, $\hat{\omega}_{ij} = \hat{P}(r_{ij} = 1|X_i, h_{ij})^{-1} = [\lambda_{i1}(\hat{\alpha}) \times \cdots \times \lambda_{ij}(\hat{\alpha})]^{-1}$, where $\lambda_{ij}(\hat{\alpha})$ is the predicted probability obtained from the logistic regression.

S2. Compute an initial estimate of β by using an ordinary generalized linear model, assuming independence of the responses.

S3. Compute the working correlation matrix $\mathbf{R}$ based on the standardized residuals, the current estimate of β and the specified structure of $\mathbf{R}$.

S4. Compute the $n_i \times n_i$ estimated covariance matrix: $\mathbf{V}_i = \phi A_i^{1/2} \hat{\mathbf{R}}_i(\alpha) A_i^{1/2}$

S5. Update $\hat{\beta}$:

$$\hat{\beta}_{r+1} = \hat{\beta}_r + \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta} \right) \mathbf{V}_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta} \right)' \right]^{-1} \left[\sum_{i=1}^{N} \frac{\partial \mu_i'}{\partial \beta} \mathbf{V}_i^{-1} W_i (Y_i - \mu_i) \right]$$

S6. Steps S3-S5 are repeated until convergence.

In SAS, to estimate the probabilities for dropout as well as to pass the weights (predicted probabilities) to be used for WGEE, the “dropout" and “dropwgt" macros introduced by [17] are used for this purpose and the macros need no modification. The variables “dropout” and “previous” are constructed through the use of the dropout macro. The dropout outcome variable is discrete indicating an individual drops out of the study before the end of follow-up, whereas, the previous variable refers to the outcome at previous occasions. Second, the dropwgt macro is used to pass the weights to the individual observations in the WGEE. Such weights are calculated as the inverse of the cumulative product of conditional probabilities, estimated as $\hat{\omega}_{ij} = 1/(\hat{\lambda}_{i1} \times ... \times \hat{\lambda}_{ij})$. This simply means the use of the predicted probabilities from the fitted missingness model to calculate the weights. Lastly, once the earlier two steps are executed appropriately, the last step is implemented by specifying the weights, by means of the weight statement in SAS procedure GENMOD. In specifying the working correlation matrix, there are a number of choices but in our case the first order autoregressive AR(1), TOEP and compound symmetry (CS) are chosen. However, there are two features to note about procedure GENMOD:

- This procedure is more appropriate for an independence working correlation matrix structure. In the GENMOD procedure, the weight statement procedure does not properly include weights when other correlation structures are used.
- The GENMOD procedure regards the weights as fixed. Consequently, the standard errors of the regression parameters from the two-step approach are conservative, which leads to wider confidence intervals and conservative inference [7].

In SAS/STAT 9.4, the above deficiencies are better handled with the new GEE procedure which also provides appropriate standard errors. Furthermore, PROC GEE also handles a variety of working correlation structures. Thus in our case, we use PROC GEE in order to exploit this flexibility

3.1.2 WGEE based on subject-specific weights

The subject-specific weighted method is quite different from the observation-specific weighted method because it assigns a single weight for all observations within an individual. This means all the observations from an individual receive the same weight. Using this technique, one obtains the regression parameter estimates from the subject-specific weighted generalized estimating equation given by

$$\mathbf{S}_{sw}(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i'}{\partial \beta} \mathbf{V}_i^{-1} w_i (Y_i - \mu_i(\beta)) = 0 \tag{4}$$

where the weight w_i for individual i happens to be the inverse probability of an individual i dropping at the observed time [8, 18]. Remember w_i is a scalar, as opposed to the weight matrix W_i in the observation-specific WGEE. The estimating equation from the subject-specific weighted method is unbiased after the observations have been weighted properly, and this produces consistent estimates for the regression parameters β.

However, the weight w_i needs to be estimated because they are unknown. Assume that m_i is a dropout indicator for the individual i, where $m_i = \sum_{j=1}^{n_i} r_{ij} + 1$. The first visit observation y_{i1} is assumed to be always observed with $r_{i1} = 1$. Thus, the values of m_i are 2,...,n_i. Note that $m_i = n + 1$ indicates that individual i completes all the n visits, which were set aprior by design.

The definition of the weight w_i is as follows: if an individual i drops out before completing the last visit (i.e. $m_i < n+1$), then $w_i = P(r_{im_i} = 0, r_{im_i-1} = 1|X_i, h_{ij})^{-1}$. Otherwise, the individual completes all the n visits (i.e. $m_i = n+1$), and $w_i = P(r_{in_i} = 1|X_i, h_{ij})^{-1}$.

As with observation-specific weights, the dropout process implies that subject-specific weights can be estimated as a cumulative product of conditional probabilities:

- $\hat{w}_i = P(r_{im_i} = 0, r_{im_i-1} = 1|X_i, Y_i)^{-1} = [\lambda_{i1}(\hat{\alpha}) \times \cdots \times \lambda_{im_i-1} \times (1 - \lambda_{im_i-1}(\hat{\alpha}))]^{-1}$, if $m_i < n + 1$
- $\hat{w}_i = P(r_{in_i} = 1|X_i, Y_i)^{-1} = [\lambda_{i1}(\hat{\alpha}) \times ... \times \lambda_{in}(\hat{\alpha}))]^{-1}$, if $m_i = n + 1$

After the estimation of λ_{ij} by using the appropriate logistic regression for the dropout process, the subject-specific weights $\hat{w}_i$ can be obtained. There is a clear similarity in the algorithm that fits the marginal models for subject-specific WGEE technique and observation-specific WGEE technique, thus the fitting algorithm is not repeated here. Thus, the same SAS macro can be adapted to fit the subject-specific WGEE model.

3.2. Multiple imputation based GEE (MI-GEE) approach

Multiple imputation is a simulation-based approach for filling in the missing values multiple times which leads to multiple complete data sets. It is assumed that the model for the vector of repeated measurements Y_i is described by the parameter vector β. In the imputation stage, the objective is to impute the missing values with draws from the conditional distribution $f(y_i^m|y_i^0, \beta)$. However, β is not known hence an estimate for it denoted by $\hat{\beta}$, has to be obtained from the data, after which $f(y_i^m|y_i^0, \hat{\beta})$ is used to fill the missing observations. In the process, it means that we generate draws from the distribution of $\hat{\beta}$, which requires that we take into account the sampling uncertainty of estimating β. Another alternative is the Bayesian method, where the uncertainty about β is incorporated by means of using some prior distribution for β. However, after the formulation of the posterior distribution of β, the following imputation algorithm can be adopted: a random $\hat{\beta}$ is drawn first from the posterior distribution of β. The posterior distribution is approximated by the normal distribution. Then a random $\tilde{Y}_i^m$ is selected from $f(y_i^m|y_i^0, \hat{\beta})$. The so-imputed missing values are next augmented to the observed data, producing complete data, $Y_i = (Y_i^0, \tilde{Y_i^m})$. These are used to obtain $\widehat{\beta}$ and its variance, $V = \hat{Var}(\hat{\beta})$. The steps mentioned above are independent and repeated a number of times, say M times, to generate $\hat{\beta}^m$ and $\hat{V}^m$, for $m = 1, \dots, M$. Moreover, the last step as stated above is when the results of the analysis from the M completed (imputed) data are combined into a single inference. The overall estimated parameter for β and V are as follows:

$$\bar{\hat{\beta}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}^m, \tag{5}$$

and

$$V = W + \left(\frac{M+1}{M}\right)B \tag{6}$$

where

$$W = \sum_{m=1}^{M}\frac{\hat{V}^m}{M} \tag{7}$$

and

$$B = \sum_{m=1}^{M}\frac{(\hat{\beta}^m - \bar{\hat{\beta}})(\hat{\beta}^m - \bar{\hat{\beta}})'}{M-1} \tag{8}$$

The within-imputation variance and between-imputation variance are denoted by W and B respectively [20]. With the description of MI above, this gives an insight to another method of handling missingness when the MAR-based MI is combined with a GEE analysis as the substantive analysis model. MAR-based MI hinges on the flexibility of the MI procedure hence the need to understand the idea on uncongenial imputation. Uncongeniality was introduced by [16] for an inconsistent imputation model in relation to the substantive analysis model. As stated by [16], one of the greatest strength of MI is that these two models (substantive and imputation) can be inconsistent in the sense that the two models need not be derived from the same overall model for the complete data. This method has become known as the MI-GEE technique, where M multiple data sets are subjected to the GEE analysis before the combination or pooling step. This serves as an alternative to likelihood and WGEE inference.

3.3. Doubly robust based GEE (DR-GEE)

The doubly robust DR method is an alternative approach that uses the inverse probability weights (IPW) to refine estimates of the model parameters [1], within a GEE analysis. In this technique, there is a requirement for the specification of two models: the first model is on the distribution of the complete data which include the outcome and covariates, and secondly a model for the missingness mechanism. The parameter estimates would be asymptotically unbiased when one of the models is correctly specified. On the other hand, the methods can be unstable in practice, especially when both models are misspecified [28], and can be disastrous when the propensity score (i.e. the probability of being observed) are close to zero [28, 30]. In current application, we combine IPW with MI and the GEE as the analysis model to construct DR-GEE. The robustness of the imputation model is enhanced by ensuring adequate information is included in the model, while avoidance of bias from the final inference is the target.

The main idea of the DR-GEE estimation; is to estimate the propensities for each incomplete variable conditional on the other variables, and impute the missing values on that variable by the inclusion of propensity functions (i.e. IPW) into the imputation model. The results of the analysis from M completed (imputed) data are combined into a single inference with the GEE. The expectation of this method is to be readily robust, and by design it is aimed at handling incomplete data with any pattern of missingness.

3.3.1 Doubly robust estimation

As a caricature of the analyses, it helps to consider the estimation of a population mean outcome in the presence of incomplete data. This problem shows fundamental challenges involving inverse probability weighting. Consider a study design that aims to obtain independent and identical distributed data. Let $\{(Y_i, X_i), i = 1, ..., n\}$, with Y_i as the outcome and X_i a set of auxiliary covariates for individual i. In the presence of missing data, the estimation of the mean $E(Y)$ is complicated by the fact that Y_i is not available for all individuals. Let R_i denote the missingness indicator, coded as $R_i = 1$ when Y_i is observed and $R_i = 0$ if Y_i is missing. The observed data can then be described as the random sample $\{Z_i(R_iY_i, R_i, X_i), i = 1, ..., n\}$ as illustrated in [29]. Assume that the covariates X_i contain sufficient information to explain missingness so that the missing at random assumption, $Y_i \perp R_i | X_i$ [27], holds. Let μ represent its (unknown) population value; in particular, $E(Y) = \mu$. When the outcome data are missing, consistent estimation of μ requires specification of at least one of the two following working models as stated by [21]. The probability of observing the data which is referred to as the propensity score (PS), is the first working model. This is taken as $P(R = 1|X) = \pi(X; \gamma)$, for which we assume $\pi(X; \gamma) > 0$ with probability one, where $\pi(X;\gamma)$ is a known function, smooth in γ which is an unknown $p-$dimensional parameter; for example, a logistic regression model $\pi(X;\gamma)$=expit($\gamma_1 + \gamma_2^T X$). This model is denoted as $M(\gamma) = \{\pi(X;\gamma) : \gamma \in \Re^P\}$. The second working model is for the conditional mean outcome $E(Y|X) = m(X;\beta)$, where $m(X;\beta)$ is a known function, smooth in β which is an unknown q-dimensional parameter; for example, a linear model $m(X;\beta) = \beta_1 + \beta_2^T X$ for a continuous outcome Y. This model is denoted as $M(\beta) = \{m(X;\beta) : \beta \in \Re^q\}$. As outlined in [25], the DR estimator of μ, with $\tilde{E}_n(U) = n^{-1}\sum_{i=1}^{n} U_i$, can be obtained as

$$\hat{\mu}_{DR}(\hat{\gamma},\hat{\beta}) = \tilde{E}_n\left[\frac{RY}{\pi(X,\hat{\gamma})} - \frac{R-\pi(X,\hat{\gamma})}{\pi(X,\hat{\gamma})}m(X,\hat{\beta})\right] \tag{9}$$

for root$-n$ consistent and asymptotically normal estimators $\hat{\gamma}$ and $\hat{\beta}$ for the parameters γ and β [27]. This estimator is consistent for μ under the union model $M(\gamma)\cup M(\beta)$ as long as one but not necessarily both working models are correctly specified. If the intersection model $M(\gamma)\cap M(\beta)$ holds, that is, both working models are correctly specified, the DR estimator in equation (9), is locally efficient [27] under model $M(\gamma)$. It then has the smallest asymptotic variance within the class of all estimators that are consistent and asymptotically normal under $M(\gamma)$, provided that also $M(\beta)$ is correctly specified, with more explanation in [29]. Note that if $R_i = 0$ in equation (9) then the contribution in the summation is $m(X_i,\hat{\beta})$. If on the other hand $R_i = 1$ and $0 < \pi(X_i,\hat{\gamma}) = u_i < 1$ such that $u_i^{-1} = K$ then the contribution in the summation is $KY_i - (K-1)m(X_i,\hat{\beta})$. However, as a caution it is important that $\pi(X_i;\gamma)$ is bounded away from zero in the sense that $\pi(X_i;\gamma) \geq \delta_0 > 0$, otherwise one may be faced with undefined terms in the summation.

4. Simulation study

4.1. Data generation

We simulated data in order to mimic the non-Gaussian longitudinal clinical trial data. In the simulation, 1000 random samples of sizes $N = 100,\ 250$ and 500 individuals were drawn. The individuals were assumed to have been assigned to two treatment arms (Higher dose=1 and Mild dose=0). The measurements were taken at four time points ($j = 1, 2, 3, 4$). Y_{ij} is the response variable measurement from individual i, at time j. The two levels of the response take the values 1 or 0 representing the event or non-event respectively. We modeled the event probability as a function of the explanatory variables. A marginal model for each binary response variable Y_{ij} is the focus and we assumed the logistic regression model. We thus generated the longitudinal binary outcome according to the following marginal model

$$logit\ P(y_{ij} = 1) = \beta_0 + \beta_1 dose_i + \beta_2 time_i + \beta_3 dose_i \times time_{ij} \tag{10}$$

where the model parameters are β, where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$. In the model, the fixed categorical effects are treatment ($dose$), time (t) and the treatment-by-time interaction ($dose \times t$). However, time was taken as a continuous variable. We fixed $\beta_0 = -1, \beta_1 = 1, \beta_2 = 0.07, \beta_3 = -1.25$. We used AR(1) as the working correlation matrix, with common correlation ρ=0.70. Dropouts were created on the complete simulated datasets using different settings of missingness rate on the response. The dropouts were imposed on the response variable Y_{ij}. For the MAR mechanism to be achieved, after simulating a data set without missing data, we adopted the following strategy. We assume that dropout can occur after the first time point. Thus in this study, four dropout patterns are possible. That is, dropout at second, third, fourth time points or no dropout. According to [24], the data generated at time j and the subsequent times were assumed to be dependent on the values of outcome measured at time $j - 1$. In our study, we retained the criterion that if the dependent variable (Y_{ij}) was positive ($i.e. Y_{ij} = 1$), then the individual dropped out at the next time point, is j+1. We generated dropouts of approximately 10%, 20% and 30% respectively. We considered a monotone missing data pattern in our simulation where the only source of dropout was an individual's withdrawal.

4.2. Measures of performance of the techniques

The performance of the different techniques were assessed using two criteria criteria: the relative bias (RB) and root mean square error (RMSE). SAS/STAT 9.4 was used to perform the statistical analyses and to produce the results. In each case, the covariance structure used in the enhanced GEE is the compound symmetry to account for correlation in the data. The performance criteria used are briefly discussed below.

4.2.1 Relative bias

The relative bias (RB) is defined as the fractional difference between the averaged estimate and the true value. It is expressed as $RB = \dfrac{\bar{\widehat{\beta}} - \beta}{\beta}$, where β is the true parameter value of interest. If number of simulations performed is represented as S then $\bar{\widehat{\beta}} = \sum_{i=1}^{s} \dfrac{\widehat{\beta}_i}{S}$. The estimate of interest within each of the $i = 1, ..., S$ simulation is $\widehat{\beta}_i$. In addition, $\bar{\widehat{\beta}}$ is simply the estimate averaged over all simulations.

4.2.2 Root mean squared error

The mean squared error (MSE) is defined as the averaged squared difference between the parameter estimates and its corresponding true value. MSE is equal to the sum of the variance and squared bias of the parameter estimates. The RMSE is defined as the square root of MSE. This is calculated as

$$RMSE = \left[(\bar{\widehat{\beta}} - \beta)^2 + Var(\bar{\widehat{\beta}}) \right]^{1/2},$$

where

$$Var(\bar{\widehat{\beta}}) = \sum_{s=1}^{S} \frac{(\widehat{\beta}_i - \bar{\widehat{\beta}})^2}{(S-1)},$$

where S is the number of replications. The importance of RMSE is that it measures the overall precision or accuracy, therefore it is used to evaluate the performance of estimation methods. In general, the more effective technique would have a smaller RMSE [11].

4.3. The analysis

In this section, we discuss the results of the simulation study that compares the three techniques, namely the WGEE, MI-GEE and DR-GEE under different dropout settings. The imputation model for the MI-GEE and DR-GEE methods are specified accordingly and the WGEE requires no imputation. We consider a correct propensity score model for DR-GEE. The simulation study also considers the correct specified model for the imputation model for both the MI-GEE and DR-GEE. The measurement at first time point is assumed to be observed for each individual. The incomplete data set were multiply imputed and analyzed by MI-GEE and DR-GEE techniques respectively. We incorporate weights to analyze the WGEE. In the case of MI-GEE, dose and response status at other time points were included as covariates in the imputation model. The logistic regression was used to estimate the propensity scores for the DR-GEE approach, which in turn were used in the imputation model. We set the number of imputations to 50.

From Table 1, under the sample size of 100; it can be observed that the relative bias was smaller under the DR-GEE method showing better asymptotically unbiased parameter estimates, except for β_1 under 30% dropout setting. It is also observed that the RMSE based on the DR-GEE was marginally smaller than the MI-GEE, but not for all under WGEE and produces highest values. However, the results obtain for the MI-GEE under the sample size of 250 performs better than DR-GEE in terms of RB and RMSE. In addition, the results obtain from the sample size 500 is similar to the results recorded for sample size of 100. But small sample produces efficient results under the DR-GEE which is better in performance than WGEE and MI-GEE. This points to the greater efficiency of the estimators of the DR-GEE method.

Table 1: Simulation study: relative bias (RB) and root mean squared error (RMSE) values for the different parameters under the three models; WGEE, MI-GEE and DR-GEE under MAR mechanism over 1000 samples: N=100, 250 and 500 individuals, for monotone dropout.

Sample	Drp	Par	WGEE		MI-GEE		DR-GEE	
			RB	RMSE	RB	RMSE	RB	RMSE
100	10%	β_0	**-2.2232**	**2.2583**	-2.0755	2.0980	-2.0742	2.0960
		β_1	0.4993	0.5331	**0.6078**	**0.6191**	0.6042	0.6154
		β_2	**1.2300**	**0.0874**	1.1671	0.0825	1.1657	0.0824
		β_3	-0.2449	0.3464	**-0.3320**	**0.4169**	-0.3300	0.4144
	20%	β_0	**-2.2705**	**2.3106**	-1.8150	1.8356	-1.8342	1.8547
		β_1	0.5176	0.5522	**0.9435**	**0.9501**	0.9400	0.9465
		β_2	**1.2800**	**0.0908**	1.0371	0.0733	1.0486	0.0741
		β_3	-0.2585	0.3387	**-0.5426**	**0.6790**	-0.5420	0.6783
	30%	β_0	**-2.4141**	**2.4506**	-1.8894	1.9057	-1.8915	1.9077
		β_1	0.3985	0.4608	1.1780	1.1828	**1.2106**	**1.2153**
		β_2	**1.3643**	**0.0968**	-1.0943	0.0772	1.0929	0.0772
		β_3	-0.2177	0.3058	**-1.2690**	**1.5866**	-0.7485	0.9360
250	10%	β_0	**-2.4930**	**2.5272**	-1.9206	1.9434	-1.9194	1.9422
		β_1	0.5230	0.5990	0.7332	0.7431	**0.7355**	**0.7451**
		β_2	**1.3571**	**0.0963**	1.0943	0.0774	1.0943	0.0774
		β_3	-0.1506	0.2115	-0.3509	0.4404	**-0.3522**	**0.4420**
	20%	β_0	**-2.0565**	**2.0944**	-1.7304	1.7512	-1.7016	1.7224
		β_1	0.6403	0.6725	1.0594	1.0656	**1.0750**	**1.0811**
		β_2	**1.1414**	**0.0814**	1.0014	0.0709	0.9843	0.0697
		β_3	-0.2386	0.3174	-0.5541	0.6934	**-0.5682**	**0.7042**
	30%	β_0	**-2.0379**	**2.0726**	-1.7165	1.7321	-1.7261	1.7419
		β_1	0.6815	0.7194	1.3698	1.3743	**1.3741**	**1.3785**
		β_2	**1.1500**	**0.0818**	1.0000	0.0706	1.0086	0.0711
		β_3	-0.3030	0.4041	-0.7754	0.9696	**-0.7800**	**0.9755**
500	10%	β_0	**-2.0181**	**2.0611**	-1.8110	1.8369	-1.8026	1.8032
		β_1	0.4197	0.4599	**0.6426**	**0.6539**	0.6179	0.6182
		β_2	**1.1243**	**0.0804**	1.0557	0.0749	1.0514	0.0736
		β_3	-0.1707	0.2316	**-0.3710**	**0.4654**	-0.3586	0.4484
	20%	β_0	**-1.9184**	**1.9607**	-1.6048	1.6280	-1.6188	1.6416
		β_1	0.5522	0.5872	**0.9696**	**0.9764**	0.9610	0.9679
		β_2	**1.1043**	**0.0789**	0.9529	0.0676	0.9586	0.0679
		β_3	-0.2658	0.3474	**-0.5747**	**0.7191**	-0.5703	0.7136
	30%	β_0	**-2.0692**	**2.1084**	-1.7427	1.7584	-1.7605	1.7761
		β_1	0.5309	0.5770	**1.3415**	**1.3461**	1.3037	1.3085
		β_2	**1.1971**	**0.0853**	1.0171	0.0718	1.0314	0.0728
		β_3	-0.2974	0.3963	**-0.8162**	**1.0206**	-0.7973	0.9970

4.4. The application

The data used in this paper show the application of the three modifications to the GEE procedure when dealing with longitudinal data with missing observations. The data set used is from a longitudinal clinical trial study of women that used contraception during the four consecutive months [7]. Out of the 1,151 women available for the study each of them were randomly assigned to one of two treatments available: 100 mg or 150 mg of depot-medroxyprogesterone acetate (DPMA) representing the low and high dose of the drug respectively. The Amenorrhea status in each of the four months was measured as the response variable. The research question was on the effect of treatment on the rate of the Amenorrhea over time. More details about the study can be found in [7].

Let y_{ij} denote the Amenorrhea status of the ith woman at the jth visit, $j = 1, \dots, 4$, and suppose $\mu_{ij} = P(y_{ij} = 1|x_{ij})$ denote the probability of a positive Amenorrhea status at visit j to individual i given covariates information x_{ij}. In order to determine the effect of treatment on the rate of Amenorrhea over time, we consider the following marginal model:

$$\text{logit }(\mu_{ij}) = \beta_0+\beta_1 time_{ij}+\beta_2 dose_i+\beta_3 time_{ij}^2+\beta_4 dose_i \times time_{ij}+\beta_5 dose_i \times time_{ij}^2$$

Of the 1,151 women in this study, 576 are from the low-dose group, and 575 are from the high-dose group. For the low-dose group, 62.67% of the women completed the trial; for the high-dose group, 61.39% of the women completed this trial. Thus, both groups have substantial dropouts.

We considered the following logistic regression model for the missingness mechanism to obtain the weights for the wGEE:

$$\begin{aligned} \text{logit } p(r_{ij} = 1|r_{ij-1} = 1, dose_i, time_{ij}, y_{ij-1}) \\ = \alpha_0 + \alpha_1 I(time_{ij} = 2) + \alpha_2 I(time_{ij} = 3) \\ + \alpha_3 dose_i + \alpha_4 y_{ij-1} + \alpha_5 dose_i \times y_{ij-1} \end{aligned} \tag{11}$$

Equation (11), is the logistic regression for the missingness model where the second and third terms are the copy of time used as a class or factor variable. The fifth term is to relate the probability that a participant will dropout to previous Amenorrhea status. The last term relates the probability that a participant will dropout to the interaction of dose and previous Amenorrhea status.

The large fraction of missing observations pose a challenge in this trial study, where the pattern of missingness is a monotone dropout. Using standard GEE may produce biased estimates because it is near impossible to justify an MCAR assumption. Furthermore, complete case analysis would result to a heavy loss of data due to a large fraction of missing values. WGEE is possible with a monotone missingness pattern and difficult when the pattern of missingness is intermittent. Imputation strategy also gives consistent parameter estimates of interest.

The results from the three modifications of the GEE procedure are shown in Table 2. The first one is the weighted method (WGEE) using observation-specific weights model;

Table 2: Parameter estimates (Est), standard errors (SE), p-value obtained from the Amenorrhea data under the methods of (WGEE), MI-GEE and DR-GEE under MAR mechanism using different working correlation structure.

		WGEE			MI-GEE			DR-GEE		
Cor str	Par	Est	SE	$Pr > \|t\|$	Est	SE	$Pr > \|t\|$	Est	SE	$Pr > \|t\|$
	β_0	-2.2057	0.1391	<.0001	-1.9573	0.2340	<.0001	-1.7523	0.2411	<.0001
	β_1	0.3672	0.1691	0.0298	0.4578	0.1825	0.0121	0.2496	0.1971	0.2054
CS	β_2	-0.4233	0.2068	0.0407	-0.3923	0.3340	0.2401	-0.2958	0.3417	0.3867
	β_3	0.0857	0.0500	0.0868	0.0097	0.0332	0.7712	0.0125	0.0363	0.7299
	β_4	0.5850	0.2536	0.0211	0.5726	0.2607	0.0280	0.5071	0.2793	0.0189
	β_5	-0.1530	0.0743	0.0395	-0.1095	0.0473	0.0207	-0.1042	0.0512	0.0418
	β_0	-2.2039	0.1392	<.0001	-1.9502	0.2347	<.0001	-1.7435	0.2417	<.0001
	β_1	0.3659	0.1689	0.0302	0.4555	0.1827	0.0127	0.2461	0.1975	0.2126
	β_2	-0.4156	0.2064	0.0440	-0.3340	0.3329	0.3158	-0.2760	0.3427	0.4207
AR(1)	β_3	0.0860	0.0501	0.0858	0.0097	0.0332	0.7746	0.0123	0.0363	0.7340
	β_4	0.5851	0.2527	0.0206	0.5596	0.2600	0.0314	0.5043	0.2795	0.0712
	β_5	-0.1547	0.0743	0.0374	-0.1078	0.0471	0.0220	-0.1036	0.0510	0.0422
	β_0	-2.2012	0.1418	<.0001	-1.9425	0.2362	<.0001	-1.7270	0.2439	<.0001
	β_1	0.3644	0.1687	0.0308	0.4532	0.1830	0.0133	0.2404	0.1979	0.2243
TOEP	β_2	-0.4004	0.2087	0.0551	-0.3279	0.3351	0.3278	-0.2790	0.3469	0.4212
	β_3	0.0861	0.0501	0.0855	0.0093	0.0331	0.7791	0.0121	0.0362	0.7378
	β_4	0.5809	0.2512	0.0207	0.5571	0.2605	0.0325	0.5051	0.2805	0.0718
	β_5	-0.1565	0.0743	0.0391	-0.1078	0.0469	0.0217	-0.1035	0.0509	0.0419

Notes: The missing value is on the response variable and are approximately 62.67% and 61.39% on low-dose and high-dose groups respectively.

the second is the multiple imputation using multiple imputation in SAS/STAT 9.4 before GEE; and the third is the doubly robust technique using inverse probability weighting and imputation models, respectively. Three different working correlation structures were adopted. However, we briefly explain the result obtain using compound symmetry (CS) because the result is similar to other results. Thereafter, we compare the results of the three methods used. Furthermore, it is also noted that the p-value for β_3 i.e quadratic time effect is not significant under all the three techniques. All the techniques provided the same conclusion for the effect of dose (β_2). The negative effect of dose indicates that the rate of change of log odds Amenorrhea over time is lower in the group receiving 150 mg compared to the reference group receiving 100 mg of depot-medroxyprogesterone acetate (DMPA). Then β_5 i.e. dose and quadratic time effect shows negative effect which indicates that the rate of change of log odds Amenorrhea over quadratic time depends on the dose, and is non-linear.

For the purpose of comparison, under the WGEE TOEP produces the lowest parameter estimates except for β_3, β_4 and β_5 under the CS. In the case of MI-GEE, TOEP records the best values without any exceptions, but under the DR-GEE the situation is different as TOEP gives parameter estimates that are smaller than other methods, except for AR(1) where β_2 and β_4 produce the lowest values. Furthermore, the standard errors produced are small and closer to one another. This study has given an insight that the use of an appropriate correlation structure could produce better parameter estimates.

These results show that the two techniques perform better than the WGEE. Combining these results and the relative performance for the simulation study suggests that both MI-GEE and DR-GEE which are imputation based are quite strong methods. The WGEE had smaller standard errors but this did not change the overall inference and conclusion. The appealing feature in using the DR-GEE method is its doubly robust property.

5. Discussion and conclusion

In this paper, the focus was on the performance of three different techniques for handling longitudinal binary data, under the MAR assumption with monotone dropout as the pattern of missingness. In addition, we prioritize the use of different working correlation structures to find out whether it would affect the parameter estimates and standard errors substantially. Therefore, we presented three stand alone enhancements to the generalized estimating equations for incomplete binary longitudinal data under MAR. Three methodologies were used namely multiple imputation, inverse probability weighting and its doubly robustness counterpart. The main focus was on DR-GEE technique for handling incomplete binary measurement because it combines both the weighting and imputation remedies to handle incompleteness. Furthermore, another attraction to this method is that it needs only the correct specification of at least one of the models, but not not necessarily the two. However, in the simulation results when one of the missingness or outcome models is correct; the doubly robust estimators are consistent and present small-sample bias when compare with the single robust alternatives WGEE and MI-GEE.

But DR-GEE has smallest standard errors than WGEE and MI-GEE especially when the sample size is small. In real application, the predictive model was not misspecified and this made the doubly estimators had a great potential of reducing the bias when the MAR assumption is correct.

In our study, we adopted different working correlation structures and observed differentials in the parameter estimates under the different methods used. We observed smaller estimates under TOEP than we have under AR(1) and CS which is an indication that parameter estimates under TOEP consistent and better that AR(1) ans CS. On the comparison between WGEE and MI-GEE [2, 4] among others provided evidence of preference of MI-GEE over WGEE in longitudinal binary data. In addition, in the study conducted in [17] assumed independent working correlation structure,but in our study different correlation structures were used which serve as an extensions.

References

[1] Bang H and Robins, JM. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrika* 61(4):962–973.

[2] Beunckens C, Sotto C and Molenbergs G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal Data. *Computation Statistics and Data Analysis* Vol.52:3 pp 1533–1548.

[3] Carpenter JR, Kenward MG and Vansteelandt S. (2006). A Comparison of Multiple Imputation and Doubly Robust Estimation for Analyses with Missing Data. *Journal of the Royal Statistics Society* Series A-Statistics in Society 169(3): pp 571–584.

[4] Clayton D, Spiegelhalter D, Dunn G and Pickles A. (1998). Analysis of Longitudinal Binary Data from Multiphase Sampling. *Journal of the Royal Statiatics Society Statistical Methodology*: Series B 60:1 pp 71–87.

[5] Donneau AF, Mauer M, Molenbergs G and Albert A. (2014). A Simulation Study comapring Multiple Imputation Methods for Incomplete Longitudinal Ordinal Data. *Communications in Statistics - Simulation and Computation* 44:5, 1311–1338. DOI:10/1080/03610918.2013.818690

[6] Donneau AF, Mauer M, Murielle, Lambert P, Molenberghs G and Albert A. (2015). Simulation-based study comparing Multiple Imputation Methods for non-monotone ordinal data in longitudinal settings. *Journal of Biopharmaceutical Statistics* 25(3):570–601.

[7] Fitzmaurice GM, Laird NM and Ware JH. (2011). *Applied Longitudinal Analysis.* Hoboken, NJ: John Wiley and Sons.

[8] Fitzmaurice GM, Molenberghs G and Lipsitz SR. (1995). Regression Model for Longitudinal Binary Responses with Informative Drop-Outs. *Journal of the Royal Statistics Society*, Series B (Methodological) 57:4 pp 691–704.

[9] Funk MF et. al. Doubly Robust Estimation of Causal Effects. (2011). *American Journal of Epidemiology*. Vol 173:7 pp 761–767.

[10] Guixian L and Rodriguez RN. (2015). Weighted Methods for Analyzing Missing Data with the GEE Procedure. Paper SAS166-2015. SAS Institute Inc. pp 1–8,

[11] Huang R and Carriere KC. (2006). Comparison of Methods for Incomplete Repeated Measures Data Analysis in Small Samples. *Journal of Statistics Planning and Inference*, 136:1 pp 235–247. http://dx.doi.org/10.1016/j.jspi.2004.06.005

[12] Jolani S, Frank LE and van Buuren S. (2014). Dual Imputation Model for Incomplete Longitudinal Data. *Br J Math Stat Psychol* Vol. 67, pp 197–212.

[13] Jose LP, Enrico AC and Fabio ND. (2015). Doubly Robust-Based Generalized Estimating Equations for the Analysis of Longitudinal Ordinal Missing Data. *https://arxiv.org/pdf/1506.04451*. pp 1–17.

[14] Liang KY and Zeger SL. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* 73:1 pp 13–22.

[15] Mehrotra DV, Li X, Liu J and Lu K. (2012). Analysis of Longitudinal Clinical Trial with Missing Data Using Multiple Imputation in Conjunction with Robust Regression. *Biometrics* Vol.68:4, pp 1250–1259.

[16] Meng X-L.(1994). Multiple Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 9:4 pp 538–558.

[17] Molenberghs G, Verbeke G. (2005). Models for Discrete Longitudinal Data. *Springer Series in Statistics*, New York.

[18] Preisser JS, Lohman KK and Rathouz PJ. (2002). Performance of Weighted Estimating Equations for Longitudinal Binary Data with Drop-Outs Missing at Random. *Statistics in Medicine* 21:3035–3054.

[19] Rubin DB. (1976). Inference and Missing Data. *Biometrika* 63:3 pp 581–592.

[20] Rubin DB and Little RJ. (2002). *Statistical Analysis with Missing Data*, Second Edition, Wiley, New York

[21] Robins JM, Rotnitzky A and Zhao L. (1994). Estimation of Regression-coefficients When Some Regressors are Not Always Observed. *Journal of the American Statistical Association*, 89:427, pp 846–866.

[22] Robins JM, Rotnitzky A and Zhao LP. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence on Missing Data. *Journal of the American Statistical Association* 90:429, pp 106–129.

[23] Robins JM and Rotnitzky A. (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association* 90:429, pp 122–129.

[24] Satty A, Mwambi H and Molenberghs G. (2015). Different Methods for Handling Incomplete Longitudinal Binary Outcome due to Missing At Random Dropout. *Statistical Methodology* 24, pp 12–27.

[25] Scharfstein DO, Rotnitzky A, and Robins, JM. (1999a). Adjusting for Nonignorable Drop-Out using Semiparametric Nonresponse Models (with comments). *Journal of the American Statistical Association,* 94:448 pp 1096–1120.

[26] Toledano AY and Gatsonis, C. (1999). Generalized Estimating Equations for Ordinal Categorical Data: Arbitrary Patterns of Missing Responses and Missingness in a Key Covariates. *Biometrics,* 55(2), 488–496.

[27] Tsiatis AA. (2006). Semiparametric Theory and Missing Data, New York: *Springer Series in Statistics* pp 1024–1025, New York.

[28] Tsiatis AA and Davidian M. (2007). Comment: Demystifying Double Robusteness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science* 22(4), pp 569–573.

[29] Vermeulen K and Vansteelandt S. (2015). Bias-Reduced Doubly Robust Estimation. *Journal of the American Statistical Association,* 110:511, pp 1024–1036, DOI:10.1080/01621459.2014.958155

[30] Vansteelandt S, Carpenter J and Kenward MG. (2010). Analysis of Incomplete Data using Inverse Probability Weighting and Doubly Robust Estimators. *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences* Vol.6(1): pp 37–48.